Neurocomputing 175 (2016) 216–222

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Online robust principal component analysis via truncated nuclear norm regularization

Bin Hong, Long Wei, Yao Hu, Deng Cai *, Xiaofei He

State Key Lab of CAD&CG, Zhejiang University, No. 388 Yuhangtang Road, Hangzhou 310058, China

A R T I C L E I N F O

Article history:
Received 13 April 2015
Received in revised form
12 October 2015
Accepted 20 October 2015
Communicated by Haowei Liu
Available online 30 October 2015

Keywords:
Low rank
Truncated nuclear norm
Online learning
Robust PCA

A B S T R A C T

Robust principal component analysis (RPCA) has been widely used to deal with high dimensional noisy data in many applications. Traditional RPCA approaches consider all the samples to recover the low dimensional subspace in a batch manner, which incur expensive storage cost and fail to update the low dimensional subspace efficiently for stream data. Thus it is urgent to develop online RPCA methods. In this paper, we propose a novel online RPCA algorithm by adopting a recently proposed truncated nuclear norm as a tighter approximation of low rank constraint. Then we decompose the objective function as a summation of sample-wise cost. And we design an efficient alternating optimization algorithm in an online manner. Experimental results show that our proposed method can achieve more accurate low dimensional subspace estimation performance compared with state-of-the-art online RPCA algorithms.

1. Introduction

In many machine learning and data mining problems, one often encounters high-dimensional samples, which may contain noise caused by corruptions or outliers [26]. To recover the intrinsic low-dimensional subspace from the whole sample set, robust principal component analysis (RPCA) is intensively studied in recent years [20,3,17,2,23,21,5,25] and widely applied such as video surveillance [22], image alignment [15], document corpus modeling [14] and audio processing [10], to name a few.

In principle, a typical RPCA method assumes that the sample set can be divided into a low-dimensional component and a sparse noise component. Formally, given a sample matrix $Z \in \mathbb{R}^{m \times n}$, RPCA attempts to decompose Z as the summation of a low rank matrix X and a sparse matrix E:

$$\min_{X,E} \quad \operatorname{rank}(X) + \lambda \|E\|_0, \qquad \text{s.t.} \quad Z = X + E, \tag{1}$$

where λ is a regularization parameter.

It has been shown that the low dimensional subspace can be accurately and efficiently recovered from corrupted samples under suitable conditions [22,2]. However, the problem (1) is highly non-convex and intractable due to the rank function and l_0 norm [16]. Most researchers are seeking appropriate surrogates of the rank function and the l_0 norm and then transforming the problem (1) to a convex optimization problem [3,12,21,17]. Among them, Lin et al. applied augmented Lagrange multipliers to solve the obtained convex problem [12]. Shang et al. and Tao et al. further considered a more general case where the observation is missing and grossly corrupted, and proposed a unified framework combining RPCA and matrix competition models [17,21].

All the above works tackle this problem in a batch manner. That is, all samples are involved into iteration in the process of optimization, thus subject to two aspects of limitations. On one hand, storage cost is expensive to load all the samples in memory during the optimization procedure, especially unacceptable for large scale sample set. This is becoming a great challenge in *big data* era. On the other hand, if the samples are collected in a stream way, these methods cannot efficiently update the low dimensional subspace when a new sample is coming.

To address these problems, online approaches for RPCA emerged recently. The memory cost is independent of the scale of samples and the discovered low dimensional subspace updates quickly as a sample comes. Another important advantage of online robust PCA over batch methods is that it can track the dynamic low dimensional subspace in cases where the subspace changes with time. Thus online robust PCA can be applied to handle tasks such as video surveillance with moving cameras [4]. Goes et al. extended the batch version RPCA to the stochastic setting and provided a sub-linear convergence guarantee [5], reducing storage requirement and run-time complexity significantly. He et al. proposed an online robust adaptive subspace tracking algorithm on Grassmannian manifold [7],

* Corresponding author.
E-mail addresses: hongbinzju@gmail.com (B. Hong), weilongzju@gmail.com (L. Wei), yaoohu@gmail.com (Y. Hu), dengcai@cad.zju.edu.cn (D. Cai), xiaofeihe@gmail.com (X. He).

http://dx.doi.org/10.1016/j.neucom.2015.10.052
0925-2312/

which combines the augmented Lagrangian with the classic stochastic gradient framework. Mairal proposed a more general online dictionary learning scheme for sparse coding [13] based on stochastic approximation. Inspired by this, Feng et al. [4] and Shen et al. [18] attempt to solve RPCA problem in an online way. They adopt nuclear norm [16] and max norm [19] as surrogates of the rank function in the problem (1) respectively, then both nuclear norm and max norm can be represented in a matrix factorization form to handle sequential samples. Although convex envelopes of rank function, like nuclear norm and max norm, bring convenience in optimization procedure, they may result in nontrivial approximation error in realistic scenario [1]. Therefore, some researchers attempted to design nonconvex surrogate for rank function [20] and achieved more accurate approximation.

In this paper, we aim to address RPCA problem through an online nonconvex optimization framework. Specially, we replace the objective of minimizing matrix rank by minimizing a recently proposed truncated nuclear norm of matrix [24]. The truncated nuclear norm can also be represented in a matrix factorization form, which provides insight into estimating the incremental contribution produced by each novel sample to the whole truncated nuclear norm. Based on this, we propose an online scheme to update the low dimensional subspace as a new sample comes. Then we design an efficient iterative optimization method for implementation. Since minimization of truncated nuclear norm provides a tighter constraint to matrix rank, our algorithm can recover the low dimensional subspace more accurately. Extensive numerical results demonstrate the effectiveness of our algorithm.

The main contribution of this paper relies on the following two aspects:

- we propose an online scheme to solve RPCA problem by adopting a nonconvex surrogate of matrix rank, which is a tighter approximation compared with convex surrogates.
- we design an efficient optimization algorithm for the proposed objective function.

The rest of this paper is organized as follows: in Section 2, we introduce necessary background for our method. In Section 3, we propose the online RPCA scheme based on truncated nuclear norm. In Section 4, we design an efficient optimization algorithm to solve the objective function. Experiments are conducted to evaluate our algorithm in Section 5. Finally, conclusions are made in Section 6.

2. Preliminaries

Notations: Throughout this paper, we adopt upper case letters for matrices, e.g. X, and lower case letters for vectors, e.g. v. X_{ij} is the (i,j)th entry of a matrix X. $\|X\|_1$, $\|X\|_*$, $\|X\|_F$ stand for l_1 norm, nuclear norm, Frobenius norm of a matrix respectively. $\mathrm{Tr}(X)$ stands for the trace of a square matrix X. $\|v\|_1$ and $\|v\|_p (p \geq 0)$ stands for the l_p norm of a vector v and $\langle u, v\rangle$ stands for the inner product of vectors u and v. I stands for identical matrix.

Norm selection plays a central role in matrix related optimization problems. Although nuclear norm and max norm are two popular alternatives to substitute the low rank constraint, the truncated nuclear norm (TNN) recently proposed by Zhang et al. [24] was shown to provide a tighter and more robust approximate to the matrix rank, compared with (standard) nuclear norm, in various applications [24,9,8,27].

Given a matrix $X \in \mathbb{R}^{m\times n}$ and a nonnegative integer $s < \min(m,n)$, the truncated nuclear norm $\|X\|_s$ of X is defined as the sum of min $(m,n)-s$ minimum singular values of X, i.e., $\|X\|_s = \sum_{i=s+1}^{\min(m,n)} \sigma_i(X)$, where $\sigma_1(X) \geq \cdots \geq \sigma_{\min(m,n)}(X)$, where $\sigma_1(X), \ldots, \sigma_{\min(m,n)}(X)$ are singular values of X. In other words, $\|X\|_s$ leaves free those s largest singular values from $\|X\|_*$. The relationship between $\|X\|_s$ and $\|X\|_*$ has been illustrated in [24], as follows:

$$\|X\|_s = \|X\|_* - \max_{UU^T=I, VV^T=I} \mathrm{Tr}(UXV^T), \tag{2}$$

where $U \in \mathbb{R}^{s\times m}, V \in \mathbb{R}^{s\times n}$.

In this equation, there is no obvious relation between truncated nuclear norm and each sample, thus it is hard to estimate the contribution of each sample to the truncated nuclear norm individually. Fortunately, the nuclear norm can be factorized in the following form [19]:

$$\|X\|_* = \min_{X=LR^T} \frac{1}{2}(\|L\|_F^2 + \|R\|_F^2), \tag{3}$$

where $L \in \mathbb{R}^{m\times d}, R \in \mathbb{R}^{n\times d}$ for any $d \geq$ rank (X). Eqs. (2) and (3) suggest that $\|X\|_s$ can be formulated in the following minimization problem:

Lemma 2.1. *The truncated nuclear norm of X can be expressed as*

$$\|X\|_s = \min_{L,R,U,V} \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 - \mathrm{Tr}(ULR^TV^T),$$
$$\text{s.t.} \quad X = LR^T, UU^T = I, VV^T = I, \tag{4}$$

where $U \in \mathbb{R}^{s\times m}, V \in \mathbb{R}^{s\times n}, L \in \mathbb{R}^{m\times d}, R \in \mathbb{R}^{n\times d}$, $d \geq \mathrm{rank}(X)$.

Proof. For any U, V, L, and R satisfying $X = LR^T, UU^T = I, VV^T = I$. Eqs. (2) and (3) imply

$$\|X\|_s = \|X\|_* - \max_{U,V} \mathrm{Tr}(UXV^T)$$
$$\leq \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 - \max_{U,V} \mathrm{Tr}(UXV^T) \leq \frac{1}{2}\|L\|_F^2 + \frac{1}{2}\|R\|_F^2 - \mathrm{Tr}(UXV^T) \tag{5}$$

On the other hand, suppose the singular value decomposition of X is $X = P\Sigma Q^T$, where $P = (p_1, \ldots, p_m) \in \mathbb{R}^{m\times m}, Q = (q_1, \ldots, q_n) \in \mathbb{R}^{n\times n}$ and $\Sigma \in \mathbb{R}^{m\times n}$. Let $\hat{U} = (p_1, \ldots, p_s)^T$ and $\hat{V} = (q_1, \ldots, q_s)^T$, then [24]

$$\mathrm{Tr}(\hat{U}X\hat{V}^T) = \sum_{i=1}^{s} \sigma_i(X). \tag{6}$$

Let $\hat{L} = P\Sigma^{1/2}$ and $\hat{R} = Q\Sigma^{1/2}$, then it is straightforward to verify that $X = \hat{L}\hat{R}^T$ and

$$\|X\|_* = \tfrac{1}{2}\|\hat{L}\|_F^2 + \tfrac{1}{2}\|\hat{R}\|_F^2. \tag{7}$$

Therefore,

$$\|X\|_s = \|X\|_* - \sum_{i=1}^{s} \sigma_i(X) = \frac{1}{2}\|\hat{L}\|_F^2 + \frac{1}{2}\|\hat{R}\|_F^2 - \mathrm{Tr}(\hat{U}X\hat{V}^T). \square \tag{8}$$

This factorization provides a representation of $\|X\|_s$ based on dimensionality reduction. In this way, L can be regarded as a dictionary, then each column of R is the coefficient of a sample with respect to the dictionary L. Further decomposationdecomposition of the Frobenius norm in Eq. (4) will illustrate the explicit relationship between $\|X\|_s$ and each sample, as we present in the next section.

3. The proposed algorithm

In this section, we propose our online algorithm for the RPCA problem, which is named as *online robust principal component analysis via truncated nuclear norm regularization* (OTNNR). The main idea of our method is to take a tighter, though nonconvex, approximation to the rank operator, and then exploit an online way to solve the optimization problem. So the low dimensional